Remarks

Claims 1-13 are pending in the application. Claims 1-13 are rejected. All rejections are respectfully traversed.

The Examiner rejected claims 1-13 under 35 U.S.C. 112, first paragraph. The Examiner states that the specification does not enable "one of ordinary skill of the art to recognize the feature "video object planes."

With all due respect, the Examiner's rejection is improper. First, the application does identify "video object planes."

"Newer video coding standards, such as MPEG-4, see "Information Technology -- Generic coding of audio/visual objects," ISO/IEC FDIS 14496-2 (MPEG4 Visual), Nov. 1998, allow arbitrary-shaped objects to be encoded and decoded as separate **video object planes (VOP)**. This emerging standard is intended to enable multimedia applications, such as interactive video, where natural and synthetic materials are integrated, and where access is universal. For example, one might want to "cut-and-paste" a moving figure or object from one video to another. In this type of scenario, it is assumed that the objects in the multimedia content have been identified through some type of segmentation algorithm.

The application also goes on to site U.S. Patent Application Sn. 09/326,750 "Method for Ordering Image Spaces to Search for Object Surfaces" filed on

June 4, 1999 by Lin et al, now U.S. Patent 6,400,846. There, VOP's are sufficiently and clearly described:

"With **VOPs**, each image of a video sequence is segmented into arbitrarily shaped image regions. Each **VOP** describes a video object in terms of, for example, **shape, motion, and texture**. The exact method of producing **VOP's** from the source imagery is not defined by the standards. It is assumed that "natural" objects are represented by shape information, in addition to the usual luminance and chrominance components. Shape data can be provided as a segmentation mask, or as **a gray scale alpha plane to represent multiple overlaid objects**."

The above would make it clear to one of ordinary skill in the art what a video object plane is.

Second, under M.E.E.P. 2164.06(c) Examples of Enablement Issues, "To establish a reasonable basis for questioning the adequacy of a disclosure, the examiner must **present a factual analysis** of a disclosure to show that a person skilled in the art **would not** be able to make and use the claimed invention without resorting to **undue experimentation**. In the rejection, the Examiner does not provide the required factual analysis. The Examiner simply states that he **could not find** the "video object plane" feature in the specification. Not being able to find the claimed feature is not a factual analysis. With all due respect, the above will hopefully assist the Examiner now to find the VOP features.

Third, the above section of the M.P.E.P. with respect to the referencing prior art document states that the analysis would appear to be less critical where the features are **essentially standard components**. Certainly, the VOPs as claimed are defined in great detail in the **referenced standards** and other cited prior art documents, and would not require undue experimentation by one of ordinary skill in the art.

The invention provides a method for determining similarities of interpretation between portions of multimedia (videos) at a very high level, e.g., similar action in an adventure movie, scoring opportunities in a sports video, romantic activity in a gothic movie, fright in a horror movie, humor in a comedy movie, and so forth. The term 'high-level' is used because the similarity considers a sequence of semantic events extended over a relatively long time period.

Therefore, the invention segments multimedia content to extract video objects in the form of video object planes, which can encode arbitrary-shaped objects according to the MPEG-4 standard, also known as H.264 or AVC, see Specification, page 2:

> "Newer video coding standards, such as MPEG-4, see "Information Technology -- Generic coding of audio/visual objects," ISO/IEC FDIS 14496-2 (MPEG-4 Visual), Nov. 1998, allow arbitrary-shaped objects to be encoded and decoded as separate video object planes (VOP)... The most recent standardization effort taken on by the MPEG committee is that of MPEG-7, formally called "Multimedia Content Description Interface," see "MPEG-7 Context, Objectives and

Technical Roadmap," ISO/IEC N2729, March 1999. Essentially, this standard plans to incorporate a set of descriptors and description schemes that can be used to describe various types of multimedia content."

In the art, these newer, high-level structures are distinguished from older, low-level features such as color and motion.

Claims 1-13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Yeo et al., U.S. Patent No. 5,821,945 (Yeo), in view of Puri et al.

Video object planes (VOP) are defined in the H.264/MPEG-4 or AVC standards. The call for proposals was in May 1998, and the first draft design for the new standard was not adopted in until 1999. The Yeo patent application was filed in May 1997. Yeo could not have known about video object planes as claimed.

The invention segments multimedia content to extract video object planes. The decomposition of videos into a hierarchical scene transition graph according to Yeo reflects acts, scenes and shots of the video, not video object planes.

Yeo does not extract and associate features of the video object planes to produce content entities. Instead, the browsing process of Yeo is "automated to extract a hierarchical decomposition of a complex video selection in four steps: the identification of video shots, the clustering of video shots of

similar visual contents, the presentation of the content and structure to the users via the scene transition graph, and finally the hierarchical organization of the graph structure."

Yeo does not measure high-level temporal attributes of each content entity. Yeo states:

> "Low level vision analyses operated on video frames achieve reasonably good results for the measurement of similarity (or dissimilarity) of different shots. Similarity measures based on image attributes such as color, spatial correlation and shape can distinguish different shots to a significant degree, even when operated on much reduced images as the DC images. Both color and simple shape information are used to measure similarity of the shots."

The problems with low-level features as in Yeo are distinguished in the present application at pages 2 and 3:

> "Another problem with such low-level descriptors, in general, is that a high-level interpretation of the object or multimedia content is difficult to obtain. Hence, there is a limitation in the level of representation. To overcome the drawbacks mentioned above and obtain a higher-level of representation, one may consider more elaborate description schemes that combine several low-level descriptors. In fact, these description schemes may even contain other description schemes, see "MPEG-7 Description Schemes (v0.5)," ISO/IEC N2844, July 1999."

Yeo does not describe content entities and comparing the temporally ordered content entities in a plurality of the directed acyclic graphs to determine similar interpretations of the multimedia content.

The Examiner states that Yeo does not describe video object planes, and cites Puri. Puri cannot be combined with Yeo. The Examiner's reason that this would provide "a system for description of scene in a truly flexible manner" makes no sense. The invention is not concerned with describing scenes. The invention claims ordering multimedia content.

The invention measures attributes of content entities that include intensity attributes. The Examiner again erroneously cites Yeo, at column 7, lines 35, et seq., measures *correlations* between *frames* as differences:

Correlation of images

The inventors discovered that measuring correlation between two small images (even the DC images) does give a very good indication of similarity (it is actually dissimilarity in the definition below) in these images. By using the sum of absolute difference, the correlation between two images, f_m and f_n is commonly computed by:

$$c(m,n) = \sum_{j=1}^{J} \sum_{k=1}^{K} |f_m(j,k) - f_n(j,k)|. \qquad (8)$$

The correlation is known to be very sensitive to translation of objects in the images. However, when applied to the much reduced images, effects due to object translation appear to lessen to a great degree.

The inventors found that correlation measures can achieve clustering results as good as those done by color. To further reduce the storage space needed and increase the computational efficiency, the inventors devised a simple way to measure correlation using the luminance projection defined as follows. For a given image $f_m(j,k)$, $j=1,2, \ldots, J$ and $k=1,2, \ldots, K$, the luminance projection for the lth row is:

$$P_m^r(l) = \sum_{j=1}^{J} \mathrm{Lum}\{f_m(j,l)\}; \qquad (9)$$

and luminance projection for the lth column is:

$$P_m^c(l) = \sum_{k=1}^{K} \mathrm{Lum}\{f_m(l,k)\}; \qquad (10)$$

This is an array of size K+J and does not require the J×K storage for the whole image for correlation calculation in the later stages. To test the similarity of images f_m and f_n, the sum of absolute difference of the row and column projections is used as follows:

$$c_{lp}(m,n) = \sum_{k=1}^{K} |P_m^r(k) - P_n^r(k)| + \sum_{j=1}^{J} |P_m^c(j) - P_n^c(j)| \qquad (11)$$

The performance using the luminance projections is comparable to that of using full correlation. Thus equation (11) is used to represent the measure of correlation in two images.

There is nothing in the above about intensity attributes. Clarification is respectfully requested to exactly point out where Yeo describes measuring intensities as claimed.

The Examiner keeps reciting the same sections, yet as repeatedly point out, an extensive word search of Yeo easily reveals that the words "intensity" or "intensities" **never appear** anywhere in Yeo. The Examiner's rejection of claim 2 continues to be improper.

Yeo does not measure attributes of content entities that include direction attributes.

> Shape
>
> The present system uses as another measure of similarity between two images the two-dimensional moment invariant of the luminance. However, the inventors discovered that the order of magnitudes in different moment invariant vary greatly: in many examples the ratio of first moment invariant to the third or fourth moment invariant can vary by several orders of magnitude.

The section cited by the Examiner deals with shape.

Yeo measures, in images, the "two-dimensional moment invariant of the luminance." Those of ordinary skill in the art would not confuse **direction and luminance**, see column 7, lines 13-19. An extensive word search of Yeo easily reveals that the words "direction" **never appear** anywhere in Yeo. The Examiner's rejection of claim 3 continues to be improper.

Yeo does not measure attributes of content entities that include spatial attributes and the order is spatial. The Yeo measurements take place on frames.

Image attributes have served as the measurement of similarity between video shots at the low levels of the scene transition graph hierarchy. In tests of the present method and system, the matching of shots based on primitive visual characteristics such as color and shape resembles the process in which the users tend to classify these video shots, when they have no prior knowledge of the video sequence given to them. However, in addition to color and shape, the users are capable of recognizing the same people (in different backgrounds, clothes, and under varying lighting conditions) in different shots, and classify these shots to be in the same cluster. Further classification and grouping is possible after the users have acquired more understanding of the video sequences.

This suggests that automatic clustering schemes for the scene transition graph building can be made at multiple levels. At each level, a different criterion is imposed. The inventors considered that vision techniques are the keys in the lower levels: image attributes contribute the clustering criterion at the bottom level, image segmentation (e.g. foregrounds and backgrounds) and object recognition can be the next level of classification. In the top levels of the hierarchy, subgraph properties and temporal structures, such as discovering repeated self-loops and subgraph isomorphism, can be explored to further condense the graph.

The scene transition graph makes use of the temporal relations to mark the edges of the directed graph. Nevertheless, this structure can also be applied to the spatial relations within an image or spatio-temporal relations within a shot. In that case, one can further enhance the hierarchy of the graph to represent an even better understanding of video

Yeo does not rank order attributes measured of content entities.

The scene transition graph in Yeo is not derived form video object planes.

There is nothing in column 19 that would indicate that Yeo generates a summary of a video.

The user is given the flexibility to interactively select the number of clusters desired or to set caps on the dissimilarity values between individual shots allowed in a cluster. In test trials of the present system, after the initial shot partitions, the user only needs to slightly adjust the knobs to change these partitions to yield satisfactory results, often with less than four such trials. FIGS. **3a** and **3b** show clustering results on two sequences: a 16-minute Democratic Convention video sequence, and a News Report, respectively. The directed scene transition graph is laid out using the algorithms disclosed by László Szirmay-Kalos, in a paper "Dynamic layout algorithm to display general graphs", in *Graphics Gems IV*, pp. 505–517, Academic Press, Boston, 1994. FIGS. **4** and **5** show the sample interface and graph layout of the two above-mentioned video sequences, based on the results in FIGS. **3a** and **3b**, respectively. Each node represents a collection of shots, clustered by the method described above. For simplicity, only one frame is used to represent the collection of shots. A means is also provided for the users to re-arrange the nodes, group the nodes together to form further clusters, and ungroup some shots from a cluster. This enables the user to organize the graphs differently to get a better understanding of the overall structures.

Columns 6 though 8 also do not describe a video summary. Applicants respectfully request the Examiner to point out which word(s) in Yeo mean "a video summary." The Applicants have carefully read Yeo but cannot find any video summarization steps.

A word search reveals that the word "summary" appears exactly once in Yeo:

"SUMMARY OF THE INVENTION "

At column 7, Yeo states:

> Shape
>
> The present system uses as another measure of similarity between two images the two-dimensional moment invariant of the luminance. However, the inventors discovered that the order of magnitudes in different moment invariant vary greatly: in many examples the ratio of first moment invariant to the third or fourth moment invariant can vary by several orders of magnitude.
>
> By using the Euclidean distance of the respective moment

There is absolutely nothing there that would suggest that a measure of similarity between two-dimensional luminance would suggest a three dimensional video. A three dimensional video is a video that also includes depth, such as a MRI video or CAT scan.

Claimed are directed acyclic graphs where nodes represent the content entities and edges represent breaks in the segmentation, and the measured attributes are associated with the corresponding edges. Yeo teaches a graph "with nodes representing scenes and edges representing the progress of the story from one scene to the next."

Claimed is at least one secondary content entity associated with a particular content entity, and wherein the secondary content entity is selected during the traversing. Nowhere in columns 2-6 are these limitations described.

Claimed is a summary of the multimedia with a selected permutation of the content entities according to the associated ranks. At columns 9:

The user is given the flexibility to interactively select the number of clusters desired or to set caps on the dissimilarity values between individual shots allowed in a cluster. In test trials of the present system, after the initial shot partitions, the user only needs to slightly adjust the knobs to change these partitions to yield satisfactory results, often with less than four such trials. FIGS. **3***a* and **3***b* show clustering results on two sequences: a 16-minute Democratic Convention video sequence, and a News Report, respectively. The directed scene transition graph is laid out using the algorithms disclosed by László Szirmay-Kalos, in a paper "Dynamic layout algorithm to display general graphs", in *Graphics Gems IV*, pp. 505–517, Academic Press, Boston, 1994. FIGS. **4** and **5** show the sample interface and graph layout of the two above-mentioned video sequences, based on the results in FIGS. **3***a* and **3***b*, respectively. Each node represents a collection of shots, clustered by the method described above. For simplicity, only one frame is used to represent the collection of shots. A means is also provided for the users to re-arrange the nodes, group the nodes together to form further clusters, and ungroup some shots from a cluster. This enables the user to organize the graphs differently to get a better understanding of the overall structures.

Yeo allows the user to rearrange nodes in a graph. There is nothing there to indicate that content entities can be permuted according to *rank*.

It is believed that this application is now in condition for allowance. A notice to this effect is respectfully requested. Should further questions arise concerning this application, the Examiner is invited to call Applicants'

attorney at the number listed below. Please charge any shortage in fees due in connection with the filing of this paper to Deposit Account 50-0749.

Respectfully submitted,
Mitsubishi Electric Research Laboratories, Inc.

By
/Dirk Brinkman/

Dirk Brinkman
Attorney for the Assignee
Reg. No. 35,460

201 Broadway, 8th Floor
Cambridge, MA 02139
Telephone: (617) 621-7539
Customer No. 022199